Design for Cut Point Validation: Missouri MAP Grades 5 and 8 Science June 11–12, 2015

Ricardo Mercado, Jennifer Lord-Bessen, and Adele Brandstrom
McGraw-Hill Education CTB
March 2, 2015

Background

On June 11–12, 2015, a cut point validation workshop will be held to review cut scores for the Missouri Assessment Program (MAP) for Grades 5 and 8 Science.

CTB proposes that the Bookmark Standard Setting Procedure (BSSP; Lewis, Mitzel & Green, 1996; Lewis, Mitzel, Mercado, & Schulz, 2012) be implemented for the Grades 5 and 8 Science assessments. The implementation of the BSSP for the Missouri Science Cut Point Validation will incorporate benchmarks (Lewis, Mitzel, Mercado, & Schulz, 2012; Phillips, 2012).

During the workshop, approximately 20 Missouri educators will review the cut scores for Grades 5 and 8 Science assessments. To do so, they will use the results from the Spring 2015 online field test administration of the assessment. The implementation of the BSSP will consist of training, orientation, two rounds of judgments, and the collection of content-based rationales associated with the groups' recommendations.

Three cut scores will be reviewed for the Grades 5 and 8 Science assessments, demarcating four achievement levels: *Below Basic, Basic, Proficient*, and *Advanced*, where *Advanced* represents the highest level of knowledge, skills, and abilities.

This document describes how the BSSP will be used with workshop participants to review cut scores for the Grades 5 and 8 Science assessments for the Missouri Department of Elementary and Secondary Education (DESE). Table 1 shows a high-level overview of the workshop schedule.

Table 1. Overview of the science cut point validation workshop agenda.

Day	Time	Activities
Thursday, June 11	AM	Opening Session, Target Student Discussion
	PM	Study Ordered Item Book, Bookmark Training, Round 1
Friday, June 12	AM	Review Round 1, Round 2
	PM	Review Round 2, collect content-based rationales, evaluate workshop

History of MAP Science Standard Setting

Cut scores were established for the MAP Science tests in 2008. Under the sponsorship of DESE, CTB implemented the BSSP in Columbia, Missouri, for the tests in Grades 5, 8, and high school. The standard setting incorporated benchmarks dictated by state law at the time, which stipulated that the achievement standards must resemble those of the National Assessment of Educational Progress (NAEP). DESE worked with CTB to create ranges of bookmarks which would comply with state law: participants were free to recommend *Proficient* bookmarks within that range, thereby recommending cut scores that would comply with the law. The resulting cut scores have been used on the science assessments since 2008.

In the intervening years, small changes have been made to the tests and blueprints, and the tests have transitioned to online administration. In 2015 for the first time, students taking the online test will be administered technology-enhanced (TE) items as part of their science assessments. Although the underlying content standards for science have not changed since 2008, these incremental changes have led DESE to request that Missouri educators to validate the cut scores for Grades 5 and 8 Science.

The purpose of the cut point validation is to allow Missouri science educators to examine the tests for Grades 5 and 8 Science, to examine the existing cut scores, and to recommend whether the existing cut scores are reasonable for continued use.

Cut Point Validation Committee and Staff

The cut point validation committee will be comprised of approximately 10 participants for each grade from across the state of Missouri, for a total of 20 participants. All participants should have knowledge in the tested content. CTB recommends a majority of the committee be current Missouri teachers of science. All participants will be selected by DESE. The committee for each grade will be divided into two tables of five participants each.

Selection of Participants

The standard setting committee should represent a diverse sample of expert participants from the entire pool of all such qualified experts (Hambleton & Pitoniak, 2006). It is important that the sample is representative of the pool of experts in terms of geographic location, ethnicity, gender, community size, teaching experience, and content expertise.

All participants selected for the workshop should understand the examinee population and the content standards being assessed (Raymond & Reid, 2001; Plake, 2008). For example, teachers of students eligible for the science assessments should be included, as well as other education experts, such as school administrators or stakeholders who have an understanding of the population and assessed content, so that the decision-making process is enhanced by the diversity of the participants (Surowiecki, 2004). Some teachers should have experience teaching science to students who are gifted, Limited English Proficient, or who have disabilities.

Because the BSSP is a content-based process, it is essential that a diverse group of participants are selected and that all of the participants are experts in the content area. If a homogenous group of participants were to be selected, then the resulting cut scores could be biased, jeopardizing the validity of inferences based on the cut scores.

Prior to the workshop, DESE should assign participants to the tables such that each table has a balance of participants in terms of relevant demographic considerations (e.g., gender, geographic location). DESE should also designate one participant per table to serve as table leader. Table leaders facilitate discussions within the table groups, and are often selected teacher-educators of notable status within the committee.

Workshop Staff

Staff members from DESE and CTB will work with the cut point validation committee throughout the workshop. DESE and CTB staff members may advise the committee members on the cut point validation process, but should not influence the deliberations of the committee or make judgments that contribute to the cut score recommendations of the committee members.

Workshop Facilitators. The workshop facilitators will be Ricardo Mercado, CTB Research Manager; and Jennifer Lord-Bessen, CTB Research Scientist. Mr. Mercado and Dr. Lord-Bessen will lead the cut point validation process, including the training sessions, data collection, and data presentations. They will serve as a link between participants and DESE, as needed, and will supervise the entire workshop. They will be assisted by Adele Brandstrom, CTB Standard Setting Specialist, who will manage the flow of secure test materials and workshop data.

Group Leaders. CTB Assessment Editors will facilitate the implementation of the BSSP with the cut point validation committee. The group leader for each group will monitor the discourse at each of the tables, keep the group on schedule and facilitate discussion during the Bookmark Procedure. The group leaders are science content experts from CTB and are members of the workshop staff.

DESE Staff. DESE staff members should be present throughout the cut point validation to respond to questions from committee members. During the workshop, CTB will defer policy-related questions to DESE staff. CTB recognizes that the DESE is the sponsor of the cut point validation workshop, and is the ultimate authority over the achievement standards for the Science assessments.

Workshop Materials

The BSSP engages participants with operational test materials. Participants study these materials to operationalize the knowledge, skills, and abilities that are expected of students in each achievement level.

Ordered Item Book (OIB). The ordered item book (OIB) for each grade comprises a selection of items from the Grade 5 or 8 Science assessment. Items in the OIB will be selected from the forms selected for the 2015 operational administration. These items are ordered in the OIB in terms of difficulty. The ordering is straightforward, with easier items placed earlier in the book, followed by harder items.

Item Map. The item map summarizes the materials in the OIB. The item map indicates the order of difficulty, scale location, item number, correct response, and content standard that each item measures. On the item map, the participants answer two questions:

(1) "What does a student need to know and be able to do to respond correctly to this item?" and
(2) "Why is this item more difficult than the preceding items?"

Achievement Level Descriptors (ALDs). The achievement level descriptors (ALDs) summarize the knowledge, skills, and abilities expected of students in each achievement level. The ALDs represent the intended meaning of the achievement levels and indicate the knowledge, skills, and abilities of students who receive test scores that fall within a particular achievement level.

ALDs were established for the Grades 5 and 8 Science assessments in 2008. The ALDs, mirroring the achievement levels, were created for *Below Basic, Basic, Proficient*, and *Advanced.*

DESE has indicated that it is satisfied with the existing ALDs for Grades 5 and 8 Science, and that the ALDs will not be adjusted as part of the cut point validation. CTB recommends that participants at the cut point validation be informed of this fact, and that participants' informal suggestions of edits to the ALDs still be solicited, because these suggestions could potentially be used in future development of ALDs on other tests.

ALDs are meant to describe the intended inferences stakeholders make based upon the test scores. They serve to describe the rigor of student achievement within an achievement level, and they are used as the foundation of the participants' judgments regarding the rigor of the achievement standards.

The ALDs are used to guide discussions about the *target students* at the cut point validation. Target students are hypothetical students who possess just enough knowledge, skills, and abilities to be considered, for example, in the *Proficient* level. There are three target students for each grade: a just *Basic* student, a just *Proficient* student, and a just *Advanced* student. By discussing the target students in conjunction with the ALDs and content standards, participants at the workshop better understand the knowledge, skills, and abilities that should be held by students just at each cut score.

Target Cut Scores. The target cut scores for the workshop comprise the existing cut scores for the test. That is, the target cut scores for Grades 5 and 8 Science will be the cut scores approved by DESE following the 2008 Science standard setting.

DESE has indicated that it is satisfied with the cut scores from the 2008 standard setting. However, DESE may wish to reserve the right to adjust the target cut scores in advance of the cut point validation. For example, if the distribution of student scores is markedly different in 2014 than it was in 2008, or if the impact data in 2014 look very different than those observed in 2008, DESE may wish to adjust the cut scores to promote better articulation across grades in science. CTB will look to DESE in advance of the workshop to confirm the target cut scores for the workshop.

Workshop Agenda

The BSSP will be implemented during a two-day workshop. The workshop will comprise participant training, two rounds of the BSSP, and the documentation of participants' content-based rationales for their decisions.

Participant Training

Opening Session. The opening session is the participants' first opportunity to meet DESE and CTB staff. All cut point validation participants will be convened for a single opening session.

It is important that the participants feel appreciated and valued for their content expertise. It is also important that participants understand the purpose of the workshop, including the fact that DESE is looking to validate the cut scores on the current tests, not to change the meaning behind the achievement levels.

During the opening session, a representative of DESE will welcome participants to the cut point validation. DESE will provide an overview on the policy-based purpose of the testing program and the process that will be used to review and approve the cut scores.

Overview of the testing program. DESE should provide an overview of the goals or rationale for the current testing program as well as the evolution of the science assessments. DESE may comment on the history of the testing program, regulatory requirements, and policy decisions that are based on testing results. To highlight teacher input into the testing program, attention should be paid to any teacher-based workshops that were conducted when constructing the test. A flowchart or visual aid depicting the construction of the test is often helpful for participants.

DESE may want to pay particular attention to explaining the purpose of the testing program. Questions that participants often ask include the following:

- How have the assessments changed in recent years?
- Is the test being used to fulfill any legislative or regulatory requirements?
- How do the test results affect students and schools?
- How might the science standards be expected to change in coming years, if known?

Overview of the cut point validation process. Cut point validation is often viewed as a multi-step process in which the cut point validation workshop is one step in preserving well-articulated, reasonable achievement standards. DESE should inform participants of all the review processes that will follow their initial recommendations. It should be emphasized that the participants are making recommendations to other entities, such as DESE, that may revise these recommendations. Participants should be assured that their recommendations will be taken seriously by the entities that review the cut scores; however, the other entities may take into account additional information when making their own recommendations, such as policy-based information. A flowchart or visual aid detailing the review process can greatly clarify the process for participants.

Overview of the desired rigor of the standards. Prior to presentation of the BSSP, DESE should advise participants about the level of rigor that is expected for each achievement level. DESE should describe to

participants the overall level of knowledge, skills, and abilities that is expected of students in each achievement level.

Purpose of the workshop. DESE should make clear to participants that the purpose of the workshop is not to adjust the meaning of the achievement levels. To underscore this fact, DESE may wish to describe how the ALDs for the tests have been in use for some time, and that the ALDs will not change for the tests this year.

Instead, DESE should make it clear that the purpose of the workshop is to validate the existing cut scores on the current tests. Given the incremental changes to the science tests over time, it is reasonable that DESE periodically make sure that the cut scores still reflect the achievement levels on the test in an appropriate way. Accordingly, participants will be shown the current tests, along with information about the content standards, existing cut scores, and ALDs.

Participants will be asked to examine the tests, ALDs, content standards, and existing cut scores. If the existing cut scores are still appropriate for the test, as dictated by the content standards and ALDs, participants should recommend that they be retained. However, if the cut scores no longer appropriately reflect the expectations laid out in the ALDs, participants should recommend alternative cut scores that more closely reflect the ALDs. Participants should have content-based rationales for all of their decisions, especially if they recommend cut scores be adjusted.

Training by CTB. CTB will then provide an overview of the purpose of the cut point validation and will describe the implementation of the BSSP. Participants will be introduced to key concepts and materials of the BSSP including the ordered item book (OIB), item map, bookmarks, and target cut scores (as represented by *target OIB pages*). Participants will engage in a simulation in which they will review all of the tools of the BSSP, including a sample OIB and item map. They will simulate BSSP activities by studying the items in the sample OIB, answering questions on the sample item map, considering bookmarks, and discussing bookmark placement.

Divide into groups and tables. Participants will then leave the general training session and will assemble in their pre-assigned groups and tables. CTB recommends that DESE designate one participant at each table to be a *table leader*. Table leaders will lead a discussion among participants at their table later in the BSSP.

Orientation to computers. Each participant will be seated at a computer. Portions of the workshop will take place on the computer, and other portions through discussion with their colleagues. To begin the session within each group, group leaders will orient participants to the materials available to them on their computers, as well as the printed materials available to them.

CTB recommends that DESE advise participants on acceptable uses of the workshop materials, especially because some will be available to them electronically. For example, DESE may wish to instruct participants not to use personal electronic devices (e.g., smartphones, tablets) during the session. In addition, CTB will work to limit Internet access for participants, as appropriate, to promote electronic security during the workshop.

Study the achievement level descriptors (ALDs). Participants will be given copies of the ALDs and will be asked to study them in their tables. The group leaders will ask participants at each table to report their key findings from the ALDs.

Discuss the target students. Participants will be introduced to the three target students: the student who is just in the *Basic* level, another who is just in the *Proficient* level, and another who is just in the *Advanced* level. Participants will be asked to consider the knowledge, skills, and abilities of each of these three target students, and to take notes about these expectations.

Participants will be asked to use the target student expectations and the ALDs to support this discussion, and to focus on the knowledge, skills, and abilities that are held by students who are just barely classified in each achievement level.

Take the test. Participants will take the test to see the test from the student's perspective. To do so, participants will take an online version of one form of the student test. In doing so, participants will experience the various item types, including technology-enhanced (TE) items.

Bookmark Standard Setting Procedure

Participants will then engage in the BSSP. This proposed implementation of the BSSP incorporates benchmarks, two rounds of discussion and decision making, and the collection of participants' content-based rationales for their cut score decisions.

Presentation of the target OIB pages. Participants will be shown the target OIB pages, as based on the existing cut scores. To calculate these target OIB pages, CTB will link to the existing cut scores for each test, either through the test scale or the percentages of students classified in each achievement level (*impact data*). Each target OIB page represents a point in the OIB where a target student would be expected to have mastery of the content measured by the items prior to that page, as based on the existing cut scores. For example, if a target *Proficient* OIB page were after Page 19 of the OIB, then the just *Proficient* target student would be expected to have mastery of the content measured by the items on the first 19 pages of the OIB.

Participants will be reminded how the target OIB pages were calculated, how the existing cut scores are represented by bookmarks in the OIB, and how to consider the target OIB pages when they make their bookmark decisions. Participants will participate in training and discussion on writing content-based rationales for their bookmark placements.

Study the Ordered Item Book (OIB). On their computers, participants will study the items in the OIB in detail. As they study the items, participants will take notes on their item maps about the knowledge, skills, and abilities measured by the items.

The OIB will be presented to participants in PDF format. Participants will be able to page through the OIB on screen to study the items. If possible, participants will be asked to take notes on their item maps electronically by tiling windows on their screen for both the OIB and item map. To enhance security during this portion of the workshop, CTB recommends that Internet access be limited or blocked, so no participant could inadvertently send items using their computer.

CTB will encourage participants to discuss the items with the other participants in their tables. Participants will be encouraged to pay special attention on the items that surround each of the target OIB pages. By doing so, the participants will cultivate an understanding of the knowledge, skills, and abilities expected of students just at the cut score, and will be able to use this information as part of the cut point validation process.

Refresher training on bookmark placement. After all participants have finished examining the OIB, CTB will provide a refresher training session on bookmark placement. During this presentation, CTB will remind participants how bookmarks represent cut scores and how to use the target OIB pages to validate the cut scores. At the end of this training session, participants will be given a short quiz to test their understanding of the bookmark placement process. After this training, participants will return to their tables for Round 1.

Round 1. CTB will then instruct participants to make their Round 1 bookmark placements. CTB will instruct participants to keep the target student descriptors and target OIB pages in mind when placing their bookmarks.

Participants will be reminded that it is acceptable to recommend that the existing cut scores be retained. To do so, participants should keep the target OIB pages as their own Round 1 bookmark placements. If the existing cut scores do not appropriately reflect the ALDs on the tests, however, participants can recommend alternative cut scores by recommending different bookmarks than the target OIB pages. Participants will be encouraged to develop content-based rationales for their decisions, explicitly linking their decisions to the content-based expectations listed in the ALDs.

Participants will place their bookmarks independently. They will write individual, independent, content-based rationales for their bookmark placements.

Round 2. CTB will tabulate participants' Round 2 bookmark placements and calculate each table's cut score recommendations. A group's cut score recommendations are associated with the median bookmark placement for each cut score.

If desired by DESE, CTB will also calculate *impact data* to present to the committee. Impact data are the percentages of students who would be classified in each achievement level given a set of cut scores. CTB will describe to the group how the cut scores and impact data, if applicable, were calculated. CTB recommends that DESE be present during this presentation to respond to any policy-related questions that might arise.

Participants will discuss all bookmark placements as a grade group, across both tables. The group leaders will facilitate discussion among all participants.

After this discussion, participants will be asked to make Round 2 bookmark placements. Each participant will make recommendations: they may retain their decisions from Round 1, or they may adjust their decisions for Round 2. Again, participants will be asked to record content-based rationales for each of their decisions.

Review the Final Recommendations. Participants will be shown the median cut score recommendations from Round 2. These values represent the final recommendation of the cut score committee. If desired by DESE, participants will also be shown the impact data associated with these recommendations.

Collection of content-based rationales. Each group will then be asked to provide a single set of content-based rationales associated with its median Round 2 bookmark placements. These rationales represent the group's reasoning behind its recommendations during the cut point validation. The group leader will record the rationales for participants.

If possible, the group will recommend a single set of content-based rationales, as based on consensus within the group. If no consensus exists on the rationales, however, the group leader will elicit majority and minority rationales. All rationales will be provided to DESE alongside the group's recommended cut scores.

Evaluation. All participants will complete an evaluation of the cut point validation. Results from the evaluation will be included in the cut point validation technical report, and can be used to document participants' perception of the fairness of the BSSP and their satisfaction with the results of the cut point validation.

After the Cut Point Validation

As soon as possible after the workshop, typically within 24 hours following the workshop, CTB will share information about the participant-recommended cut scores and associated content-based rationales with DESE. This delivery is designed to help DESE and its technical advisors quickly understand participants' recommendations so the cut scores may be approved. Later, CTB will deliver two reports to DESE: a preliminary cut point validation technical report and a final cut point validation technical report.

Preliminary Cut Point Validation Technical Report

Within five working days following the end of the cut point validation, CTB will provide DESE with information on the achievement standards recommended by participants. CTB will provide DESE a preliminary cut point validation technical report containing:

- participants' recommended cut scores,
- the impact data associated with participants' recommended cut scores,
- the content-based rationales associated with participants' recommended cut scores, and
- the training materials used at the workshop.

The preliminary report is designed to be used by DESE and its technical advisors in evaluating the participants' recommended cut scores, as well as the cut point validation process as a whole.

Final Cut Point Validation Technical Report

Within 60 working days following the end of the cut point validation, CTB will prepare the final cut point validation technical report. This final report will contain the information in the preliminary report, as well as:

- narrative accounts of the events that occurred at the workshop,
- results of the participant evaluation, and
- standard error values associated the cut scores (e.g., standard errors of measurement).

The final report is designed to serve as lasting documentation of the cut point validation process. CTB will deliver this final report to DESE in electronic (PDF) form.

Cut Score Approval by DESE

CTB will look to DESE to indicate when it has approved the final cut scores for Grades 5 and 8 Science. DESE may wish to consider well in advance of the workshop the method or methods that it may use to approve the cut scores. For example, some states have taken the position that the ALDs represent the state's official policy regarding each achievement level on a test, and that the cut scores reflect a technical implementation of that policy on the assessment; accordingly, minor adjustments to cut scores without change to the ALDs do not change the official state policy, and do not necessarily require the approval of the state board. In other jurisdictions, a chief state school officer may have the authority to adjust cut scores under existing rules or regulations.

Assuming DESE will approve the cut scores in advance of the creation of the Final Cut Point Validation Technical Report, CTB will include the approved cut scores in this report. In this way, the Final Report represents lasting documentation of the cut point validation, approved cut scores, and entire process.

References

Hambleton, R. K., & Pitoniak, M. (2006). Setting performance standards. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 433–470). Westport, CT: Praeger.

Lewis, D. M., Mitzel, H. C., & Green, D. R. (1996). *Standard setting: A Bookmark approach.* Symposium presented at the Council of Chief State School Officers National Conference on Large-Scale Assessment: Phoenix, AZ.

Lewis, D. M., Mitzel, H. C., Mercado, R. L., & Schulz, E. M. (2012). The Bookmark Standard Setting Procedure. In G. J. Cizek (Ed.), *Setting performance standards: Foundations, methods, and innovations* (2nd ed., pp. 225–253). New York, NY: Routledge.

Plake, B. S. (2008). Standard setters: Stand up and take a stand! *Educational Measurement: Issues and Practice*, 27(1), 3–9.

Phillips, G.W. (2012). The benchmark method of standard setting. In G. J. Cizek (Ed.), *Setting performance standards: Foundations, methods, and innovations* (2nd ed., pp. 323–346). New York, NY: Routledge.

Raymond, M. R., & Reid, J. B. (2001). Who made thee a judge? Selecting and training participants for standard setting. In G. J. Cizek (Ed.), *Setting performance standards: Concepts, methods, and perspectives* (pp. 119–157). Mahwah, NJ: Lawrence Erlbaum.

Surowiecki, J. (2004). *The wisdom of crowds.* Boston: Little Brown.